

需求预测与分仓规划

1 赛题背景

阿里巴巴旗下电商拥有海量的买家和卖家交易场景下的数据。利用数据挖掘技术，我们对未来的商品需求量进行准确地预测，从而帮助商家自动化很多供应链过程中的决策。这些以大数据驱动的供应链能够帮助商家大幅降低运营成本，提升用户的体验，对整个电商行业的效率提升起到重要作用。这是一个困难但是非常重要的问题。我们希望通过这次的大数据竞赛中得到一些对这个问题的新颖解法，朝智能化的供应链平台方向更加迈进一步。

2 赛题介绍

高质量的商品需求预测是供应链管理的基础和核心功能。本赛题以历史一年海量买家和卖家的数据为依据，要求参赛者预测某商品在未来二周全国和区域性需求量。选手们需要用数据挖掘技术和方法精准刻画商品需求的变动规律，对未来的全国和区域性需求量进行预测，同时考虑到未来的不确定性对物流成本的影响，做到全局的最优化。更精确的需求预测，能够大大地优化运营成本，降低收货时效，提升整个社会的供应链物流效率。

3 评测指标

在本赛题中，参赛者需要提供对于每个商品在未来两周的全国最优目标库存和分仓区域最优目标库存的预测。我们会提供每一个商品的补少成本(A)和补多成本(B)，然后根据用户预测的目标库存值跟实际的需求的差异来计算总的成本。参赛者的目标是让总的成本最低。

我们定义以下变量：

- T_i : 商品 i 的全国目标库存（参赛者提供）
- T_{ia} : 商品 i 在分仓区域 a 的目标库存（参赛者提供）
- D_i : 商品 i 的未来全国实际销量（不提供给参赛者）
- D_{ia} : 商品 i 的未来在分仓区域 a 的实际销量（不提供给参赛者）
- A_i : 商品 i 的全国补少货的成本
- A_{ia} : 商品 i 在分仓区域 a 的补少货的成本
- B_i : 商品 i 的全国补多货的成本

- B_{ia} : 商品 i 在分仓区域 a 的补多货的成本

全国范围内的成本计算如下:

$$C_N = \sum_i [A_i * \max(D_i - T_i, 0) + B_i * \max(T_i - D_i, 0)]$$

分仓区域内的成本计算如下:

$$C_R = \sum_{ia} [A_{ia} * \max(D_{ia} - T_{ia}, 0) + B_i * \max(T_{ia} - D_{ia}, 0)]$$

总的衡量标准是上面两者的相加:

$$C = C_N + C_R$$

请注意我们这里需要预测的是未来两周的“非聚划算”销量，即去掉了商品参加聚划算产生的销量。我们提供的数据经过了脱敏，和实际商品的销量、浏览量和成本等有一些差距，但是不会影响这个问题的可解性。

4 数据说明

赛题数据: 我们提供商品从2014-10-10到2015-12-13的全国和区域分仓数据。参赛者需给出后面两周（2015-12-14到2015-12-27）的全国和区域分仓目标库存。

商品在各个分仓的特征包括商品的本身的一些分类: 类目、品牌等，还有历史的一些用户行为特征: 浏览人数、加购物车人数，购买人数，具体参考表1。**注意我们要预测的未来需求是“非聚划算支付件数”(qty_alipay_njhs)**，其中store_code 为“a”表示全国仓。

表 1: 商品和分仓区域粒度相关特征。

字段	类型	含义	示例	字段	类型	含义	示例
date	bigint	日期	2015-09-12	amt_alipay	Double	成交金额	0
item_id	bigint	商品ID	132	num_alipay	bigint	成交笔数	0
store_code	String	仓库CODE	1	qty_alipay	bigint	成交件数	0
cate_id	bigint	叶子类目ID	18	unum_alipay	bigint	成交人次	0
cate_level_id	bigint	大类目ID	12	ztc_pv_ipv	bigint	直通车引导浏览次数	0
brand_id	bigint	品牌ID	203	tbk_pv_ipv	bigint	淘宝客引导浏览次数	0
supplier_id	bigint	供应商ID	1976	ss_pv_ipv	bigint	搜索引导浏览次数	0
pv_ipv	bigint	浏览次数	2	jhs_pv_ipv	bigint	聚划算引导浏览次数	0
pv_uv	bigint	流量UV	2	ztc_pv_uv	bigint	直通车引导浏览人次	0
cart_ipv	bigint	被加购次数	0	tbk_pv_uv	bigint	淘宝客引导浏览人次	0
cart_uv	bigint	加购人次	0	ss_pv_uv	bigint	搜索引导浏览人次	0
collect_uv	bigint	收藏夹人次	0	jhs_pv_uv	bigint	聚划算引导浏览人次	0
num_gmv	bigint	拍下笔数	0	num_alipay_njhs	bigint	非聚划算支付笔数	0
amt_gmv	Double	拍下金额	0	amt_alipay_njhs	Double	非聚划算支付金额	0
qty_gmv	bigint	拍下件数	0	qty_alipay_njhs	bigint	非聚划算支付件数	0
unum_gmv	bigint	拍下UV	0	unum_alipay_njhs	bigint	非聚划算支付人次	0

我们还提供每个商品在全国和分仓区域的补少、补多的成本，可以用来计算总成本，具体参考表2。

表 2: 每个商品在全国和分仓区域的补少、补多的成本。

字段	类型	含义	示例
item_id	bigint	商品ID	333442
store_code	String	仓库CODE, 注意如果是全国成本，这一列是all	1
a_b	String	商品补少补多cost，用“_”链接起来。 前一个数是补少的成本，后一个是补多的成本	10.44_20.88

参赛者需要提供每个商品的全国和分仓区域的未来两周(2015-12-14到2015-12-27)目标库存，格式可以参考表3。

表 3: 每个商品在全国和分仓区域的补少、补多的成本。

字段	类型	含义	示例
item_id	bigint	商品ID	333442
store_code	String	仓库CODE, 注意如果是全国成本，这一列是all	1
target	Double	未来两周的全国或分仓区域目标库存（如果store_code为all，就是全国的target,否则就是分仓的target）	30.0

参考链接

<https://tianchi.aliyun.com/competition/information.htm?spm=5176.11165261.5678.2.21397b57ZrQAwC&raceId=231530>